Computational intelligence to simulate medical imaging

Victor Chang

Xi'an Jiaotong-Liverpool University, Suzhou, China

victorchang.research@gmail.com

Abstract

This paper describes how computational intelligence can be used to simulate medical imaging to explore areas that cannot be easily achieved by medical imaging. Simulating genes and proteins that have direct influences to cancer development and immunity belong to this category. This paper has presented simulation and inspection of BIRC3, BIRC6, CCL4 and KLKB1 with their outputs and explanations. Brain segment intensity involved with dancing has also been presented. Simulating medical imaging has been enabled by our proposed MapReduce framework with fusion algorithm, which is very similar to the digital surface theories that can best simulate how biological units can get together to form bigger units and thus eventually simulate the entire unit of biological subject. The M-Fusion and M-Update function developed by fusion algorithm can achieve a good performance evaluation that all data up to 40 GB can be processed and visualized within 600 seconds. We conclude that computational intelligence can provide effective and efficient healthcare research via simulation and visualization.

1. Introduction

Genes carry important information for each individuals and allow the parents to pass on their biological details to their children. Consequently, genetic disorders can also be passed to the next generations, or to their grandchildren. Understanding how diseases have been developed, particularly anything related to genetic disorders, or genes that are more prone to trigger other major diseases such as cancers are worth to be investigated. This may include studying malignant tumors to understand how malignant tumors can develop into cancers. To study the tumors and genes prone to tumor development, advanced methods on medical imaging and visualization will be required to allow scientists to see their objects of investigations in details, such as the abilities to magnify particular regions for investigation, simulate the biological processes of malignant tumors and help diagnose the right causes of diseases [1-3]. Modern medical imaging includes multimodality image techniques, which have become more complex and more expensive, even they are efficient. The costs of investigating patients' health have become less affordable due to the high maintenance and set-up costs [1, 4]. Hence, other pioneering techniques should be considered, such as simulations and visualization, which can replicate the real data or real area of investigations after capturing all the data and completing medical imaging successfully once. Simulations and visualizations by computational intelligence can be used at any time, results of analysis can be reproduced effectively with lower costs [5-6]. Thus, the use of advanced computational intelligence can foster a simulation-type of medical imaging, in order to fulfil concerns for both costs and quality of investigations.

The breakdown of this paper is as follows. Section 2 presents the related literature including our MapReduce Framework with fusion algorithm. Section 3 present various simulations for medical imaging including genes that can be prone to cancer development, or immunity development and brain imaging. Section 4 demonstrates results of our performance evaluation and Section 5 presents topics of discussion. Finally, Section 6 sums up Conclusion and Future Work.

2. Literature and Related Work

This section describes literature with three sub-sections. The first sub-section describes digital surfaces and development that can best be simulated. The second and third sub-sections explain our computational intelligence by our MapReduce framework and fusion algorithm.

2.1 Digital surfaces and development into tunnels, gaps and skeletons

Digital surfaces are important aspects for simulating medical images. Kim [7] define three dimensional Z cube values to present positions be adjacent to 3-cells. Movement of cells can be then defined by a mathematical framework proposed by Chen and Zhang [8]. In between cells or organs, there are tiny gaps in between, which can be called alpha-surface. Before mid-1990s, obtaining high resolution on medical images was challenging. The solution was to propose an alpha-surface using a three-dimensional R cube values to present digitization of surfaces. There are still tiny gaps while adopting both Z cube and R cube values. In order to simulate surfaces that are adjacent and tied to each other, Malgouyres [9] propose how to make Z and R cube values together by having 26 pairs of Z and R cubes adjacent to each other. Eventually, Bertrand and Malgouyres [10] develop a method that can join 26 pairs of Z and R cubes adjacent to each other better, which then forms a unit that can join other units more easily. Brimkov and Klette [11] can make this proposal a digitization successfully.

All the work described above have paved ways for developing tunnels, gaps and eventually skeletons for digital and computational medical imaging. There are gaps between different units of cells, genes, organs and biological subjects of investigations. Tunnels are in between two gaps that can form a pathway, or a channel to another biological subject of investigation. It can be between two different biological units, or two different clusters of biological units. Gaps and tunnels are important elements in simulating medical imaging, since not all details (which need powerful electron microscope) can be 100% reproduced. However, if ensuring gaps and tunnels are properly connected to the right biological units, getting close to 100% accuracy will be more likely [12]. Different biological units can form together as an integrated unit, which also represent the topological features for biological units. By forming individual biological units into an integrated or more structured unit, it allows scientists to be more confident to simulate medical imaging [13-14]. It is also a technique used to simulate a single unit of medical imaging object, before connecting all different units together to form the entire object of medical imaging simulation.

Figure 1 shows the representation of biological units, gaps, tunnels and skeletons. Each cell is a biological unit, which can form together into larger units. Spaces in between are gaps. While there are more units getting together (before forming into skeletons), tunnels can lead to different units, and any smart ways to tie all different together, can be a vital process to form into skeletons. Such “smart ways” require computational intelligence to organize and make the structure of the skeletons as tidy and coherent as possible. Each skeleton may have different shapes and size due to the subject of biological studies.

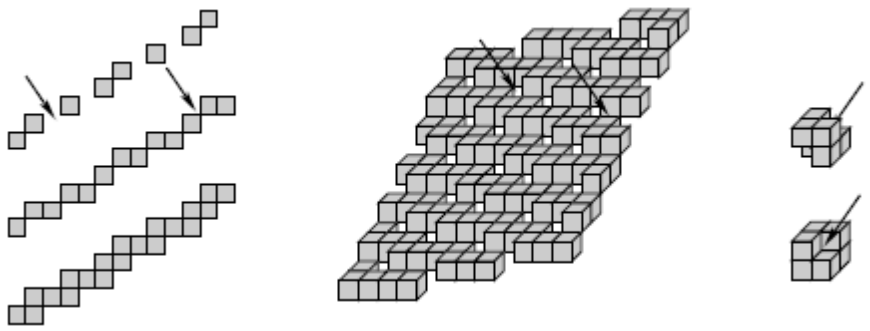


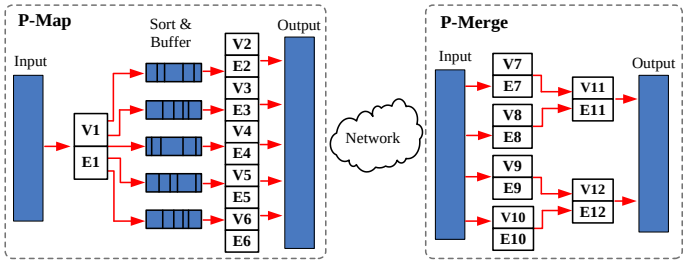
Figure 1: Illustration of biological units, gaps, tunnels and skeletons

2.2 Computational intelligence by MapReduce

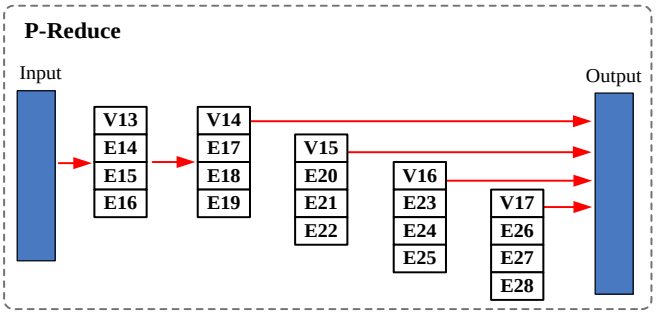
Following Section 2.1 that each biological unit can join together, a very similar concept can be applied - each biological unit can be regarded as each data to be processed, analyzed and integrated together to represent the concept of fusing all biological units together as single skeleton, or groups of skeletons. Each data carries information about the individual biological units, hence, this will require computational intelligence such as MapReduce to process all data that can be mapped, merged, reduced and then integrated together. MapReduce framework can equally split all data into independent chunks that can be processed by map function. Subsequently, all the semi-processed data can be merged together to become the input of reduce function, which will categorize all processed data and get all categorized data together as the output [15]. This process is similar to the formation of a skeleton on the left side of Figure 1.

To facilitate a smooth and fast big data processing, improved MapReduce functions have been used. Partitioning the three functions means Partitioning-Map (P-Map), Partitioning-Merge (P-Merge) and Partitioning-Reduce (R-Reduce) can focus on its tasks. MapReduce framework can present inputs as **<key, value>,** which can be summed up as <Ky, Vy>, in which y is the iteration number of MapReduce service, which can get <Ky, Vy> in each parallel pipeline. Each pipeline can determine the iteration number of inputs and outputs. By using partitioning approach, it can perform optimization of data processing and each task is divided as follows.

- P-Map: which maps all data into the grouped clusters ready processing;
- P-Merge: All the processed outputs are collected;
- P-Reduce: All the outputs can be condensed into one or fewer clusters;
- P-Query: It can be executed once all the three functions that have been executed once. The command can directly retrieve data.



(a) Illustrations of *P-Map* and *P-Merge* operations



(b) Illustrations of *P-Reduce* operations

Figure 2. Illustration of the optimization algorithm through P-Map, P-Merge and P-Reduce

Fig 2 (a) and (b) shows the illustration of our proposed optimization algorithm. *P-Map* will split all the inputs from *V* and *E* (from our Procedure 1, Partitioning algorithm) into both sorting and splitting

process. The outputs for *V* and *E* in *P-Map* reshuffled the outcomes of *V* and *E* sequentially into pairs (e.g., V_2 and E_2 and V_3 and E_3) until the end of the processing. The maximum number of pairs in each processing operation can reach up to 256 pairs. Subsequently, the output of *P-Map* will become the input of *P-Merge*, and same or identified groups of outputs are classified together. For illustration, the inputs of *P-Merge* become {V_7, E_7} all the ways to{V_{10}, E_{10}}, all of which are further processed and summed as {V_{11}, E_{11}} and {V_{12}, E_{12}}.

2.3 Computational intelligence to support fusion algorithm

This section describes how to get all outputs and fuse them together, similar to the process of establishing "individual skeletons" and "clusters of skeletons" described in Section 2.1. When all the outputs from P-Reduce are collected, they contribute to the individual skeletons. The challenge is to collect the same category of Reduced outputs to gather them together to become clusters of skeletons. To facilitate this, fusion algorithm is developed, which adopts a multimodal node to collect all the output data that can be summed up as "M-Fusion" and "M-Update". M-Fusion is a function that can combine all the output from P-Reduce function, and allows different outputs to be collected from all the nodes, and integrate the final outputs together, similar to the formation of groups of skeletons. M-Update is the function that can update all the results and confirm the formation of groups of skeletons can take place.

***Procedure* 1:** Fusion algorithm

Input: 1. Sub P-Reduced outputs R_1, R_2,…R_m and the node demands d_1, d_2,…d_n;

2. The number of replica servers, *k*.

Output: The set of chosen multimodal node, M.

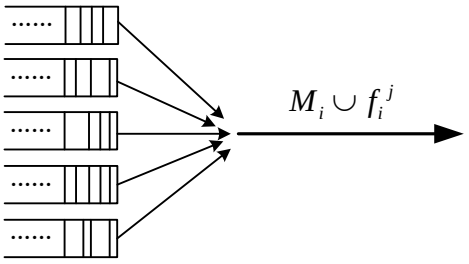
1: Read data; Map data; Merge data;

2: Reduce data; Update data;

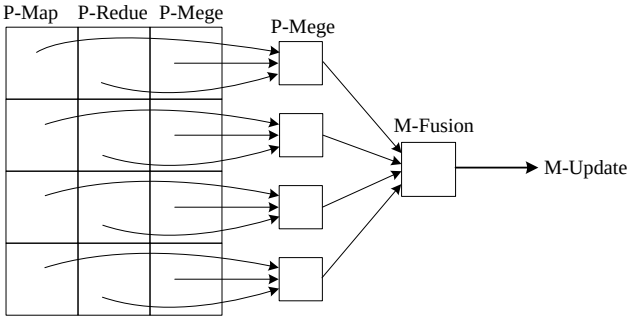
3: **Let** the node set M = Φ.
4: **For** i = 1 to k **do**
5: **For** each node f_i^j in R_i
6: **Let** f_i^j be the data fusion at each node;
7: Compute the costs of all the other nodes in C_i requesting services from node f_i^j;
8: **End for**
9: Choose the node f_i^j to achieve multi-modal data fusion at each node;
10: **Let** M = Mi $\cup f_i^j$;
11: **End for**
12: **Return** *M*
13: Fusion data; Update data; Complete multi-nodal fusion.

Figure 3 shows the Fusion algorithm receives outputs as Mi$\cup f_i^j$, indicating the multi-node and fusion numbers. M1$\cup f_1^j$ means that the output is located at the first multi-node that the fusion output is the first in the fusion service. Figure 3(a) shows the architecture, where all the outputs at each node were collected and are presented as M. At each multi-node M, a sequence number is allocated based on a first-come, first-served basis. The fusion algorithm can then map all of these nodes using the

fusion(data) and update(data) commands. For the smooth operation of these processes, the fusion commands are M-Fusion(data) and M-Update(data), as illustrated in Figure 3(b).



(a) Data fusion at multi-nodes



(b) How to achieve fusion via the five steps presented: P-Map, P-Merge, P-Reduce, M-Fusion and M-Update

Figure 3: Architecture to support fusion algorithm

For this paper, outputs of simulating medical imaging by our MapReduce framework and the fusion algorithm will be demonstrated in Section 3.

3. Simulations for medical imaging

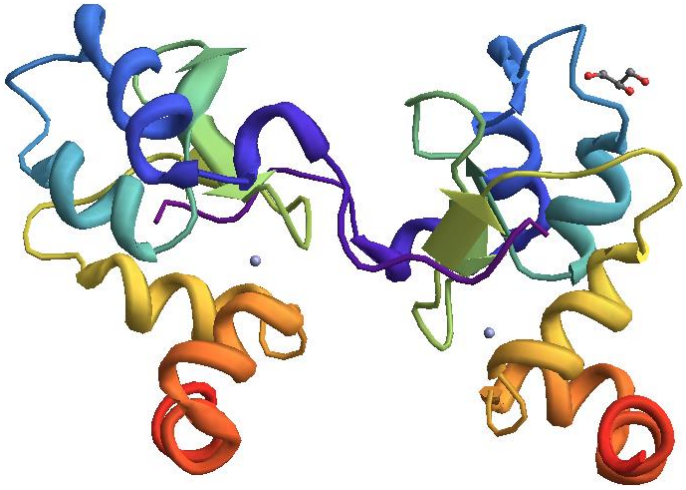
This section presents different outputs of simulating medical imaging involved with our approach to process and integrate data. For this medical research, all the cancer cells have been simulated thousands of times based on patients' different types of tumors. It has been simulated repetitively many times, so that the final outcome can be close to a better accuracy. As a result, all these thousands of simulations can become very valuable training data as described in this section for simulating high-performance malignant tumors. Genes that are more likely trigger this tumor development process, will be difficult to be detected by medical imaging. To facilitate it, first, blood sample was taken and investigated under electron microscope. Then samples of genes can be taken. Then the images can be taken. Simulation can take the size, shape and any special observations of the genes. Similarly, gene database can be cross checked to ensure that the simulated medical images look the same, except individual marks (such as cancerous spots, if there are) can be different to the healthy gene [16].

High performance and accuracy contribute to computational demands. To facilitate this, modern computing infrastructure in [15] has been built and used to simulate complex biological functions and growth. The resources include the high-end Cloud Computing, with 300 GHz multi-core CPUs, 10 Bps network speed, full virtualization Cloud Virtual Machine (VM) Center, up to 100 nodes per VM to be used for testing and simulation, previously it was used for weather simulations and forecasting [15].

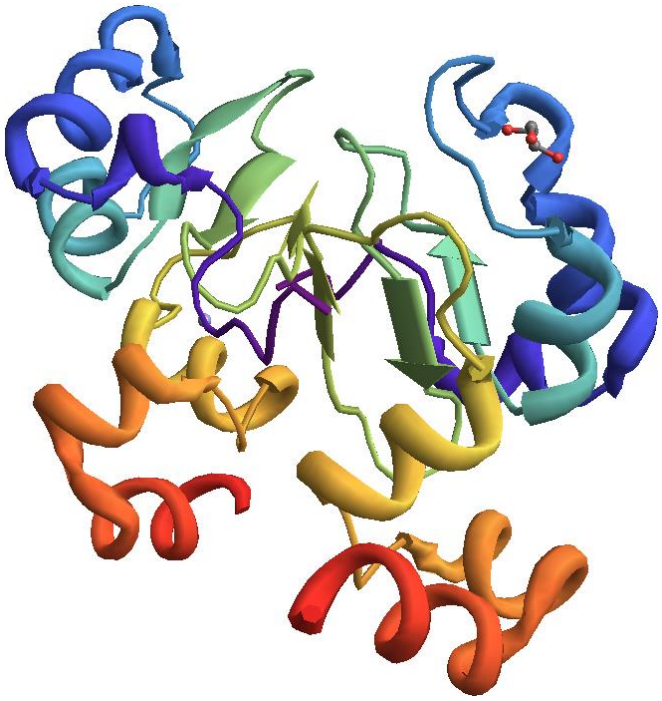
3.1 Simulation of genes as a good alternative to real medical imaging

This section describes the simulation of genes as an alternative to medical imaging following methods proposed in Section 2. First, BIRC3 is a gene that has direct impact on breast cancer and is also related

to colon cancer [17]. Figure 4 (a) shows BIRC3, which is located on chromosome 13q21. Simulating BRCA3 allows scientists to inspect the generic samples or anonymized samples from patients. Figure 4 (b) shows BIRC3 with 90 degrees of rotation to inspect any abnormities.



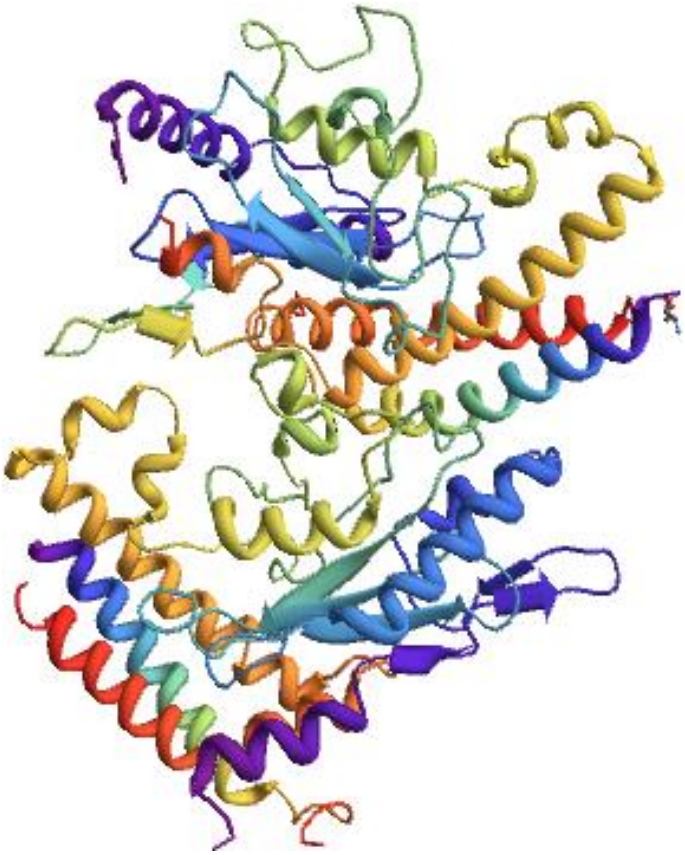
4(a): BIRC3



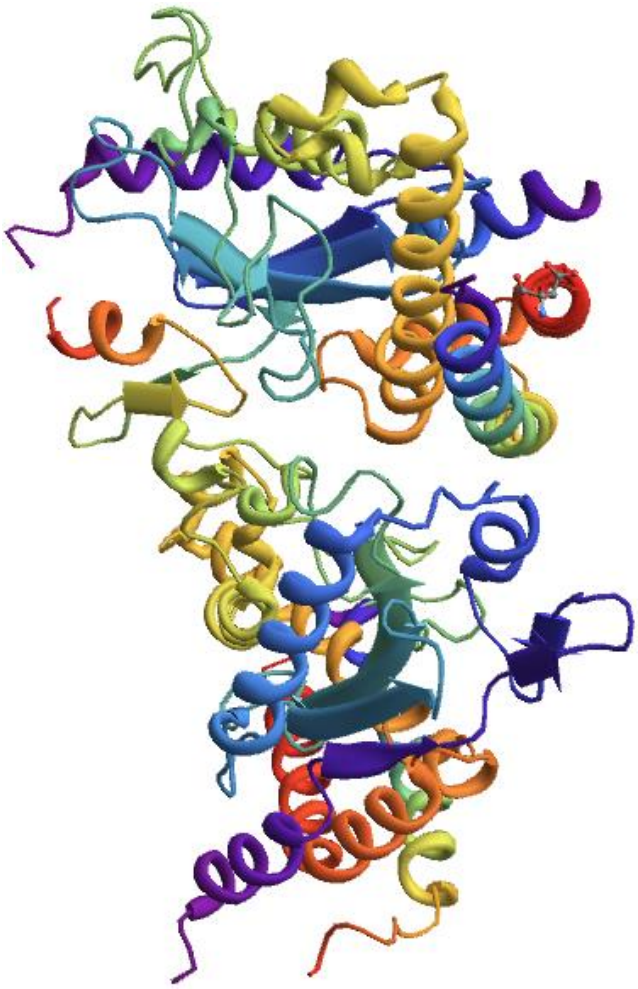
4(b): BIRC3 with 90 degrees of rotation

Figure 4: BIRC3 simulation

BIRC6 is a gene that has direct influences on brain, breast, colon and other cancers [18]. It has a more complex structure than BIRC 3. Figure 5(a) shows the BIRC6 default simulation and Figure 5(b) shows its 90 degrees of rotation. BIRC6 is worth to be investigated, since any abnormalities can be related to possibilities of having cancerous activities.



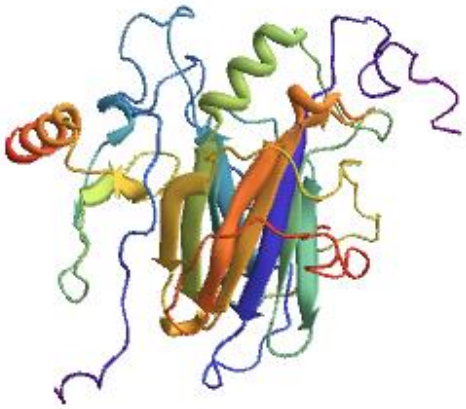
5(a): BIRC6



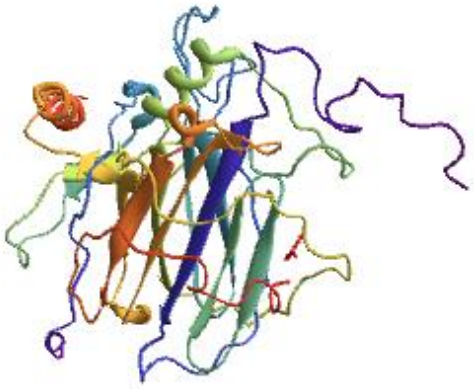
5(b): BIRC6 with 90 degrees of rotation

Figure 5: BIRC6 simulation

CCL4 is a protein that can attract natural killer cells and other immunity cells [19]. It can be regarded as an alarm to trigger the body immunity to destroy bacteria and thus its presence is useful to trigger immunity system maintenance. Checking its status can help understand how immunity can respond to cancer presence. Figure 6(a) and 6(b) show CCL4 simulation and its 90 degrees of rotation.



6(a): CCL4



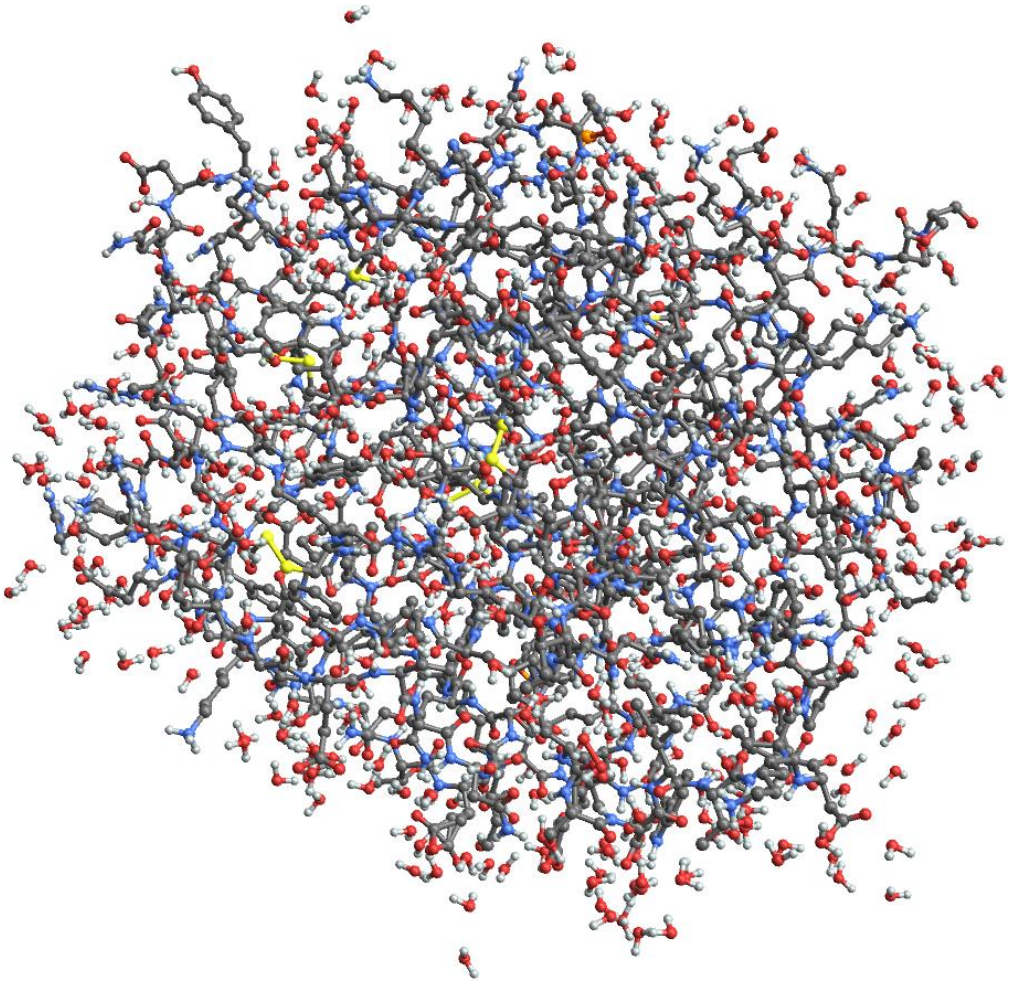
6(b): CCL4 in 90 degrees of rotation

Figure 6: CCL4 simulation

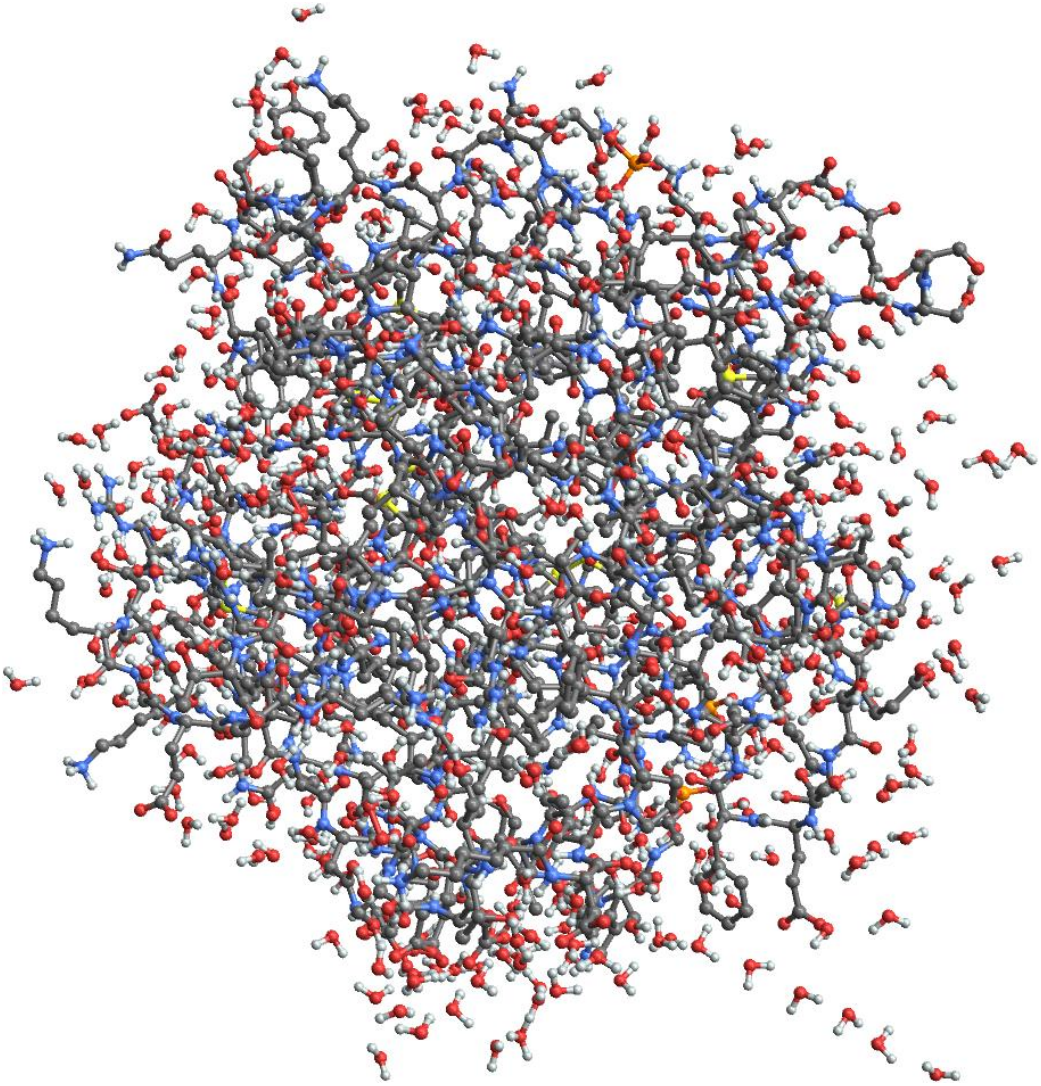
3.2 Simulating blood clot as a better alternative than medical imaging

There are other body functions that cannot be directly taken by medical imaging, such as how body immunity system can be triggered. One of these examples include blood clot simulation that at the very beginning, a gene called KLKB4 can trigger this process to happen. KLKB1 can trigger our body to make proteins called plasma pekallikrein [20]. Figure 7(a) shows KLKB1 in actions to gather all proteins in the blood as soon as possible to stop more blood flowing. Figure 7(b) shows KLKB1 in 90 degrees of rotation. Compared to Section 3.1, the difference is that in KLKB1, computational intelligence plays a more vital role. The yellow protein-like structure is the one to connect all different units, and then form into a unit of skeleton. This is also enabled by the functions of M-Fusion and M-Update to gather

all different units together from data processing and structuring. The advantage of adopting simulation approach is clear and self-explanatory since this is less practical to be achieved by medical imaging.



7(a): KLKB1 in action



7(b): KLKB1 in 90 degrees of rotation

3.3 Brain imagining

Brain imaging is often used by hospitals and medical research with different purposes such as studying the activities of the brain segments and cells, or understanding which part of the brain is active under certain types of instructions. This also includes brain segmentation and the intensity of activities on each brain segment. Figure 8 (a) shows a diagram of brain segmentation while volunteers had undergone dancing as the method to understand the intensity of brain activities on each segment. The lower part of the brain corresponds to the balance and co-ordinating body movement. Figure 8(b) shows the intensity of brain activities when volunteers are fully engaged with dancing, when their emotions are high and new dancing moves have been learned and fully adopted by the entire brain. This is the result of the collective intelligence from dancing volunteers. The way to handle data processing and analysis is based on Section 2, where each active state represents a red color dot. The intensity of redness means that segment is in an active state. This is similar to concept explained in Section 2, all data processing and fusion have taken place at the active segments.

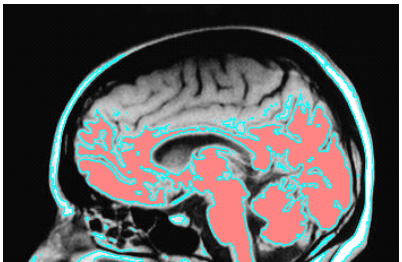


Figure 8(a): Intensity of brain activities when involved with dancing

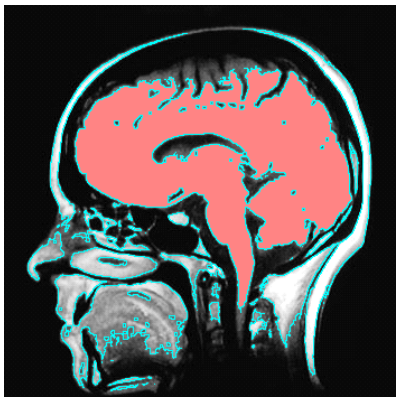


Figure 8(b): Intensity of brain activities when volunteers are fully engaged

4. Performance evaluation

Performance evaluation is a crucial part to identify the effectiveness of the proposed computational intelligence and test resiliency of medical imaging simulation, to see whether simulation can take place smoothly while a large size of data has been processed and analysed. Section 2 has described the techniques behind processing data that represent "skeleton", the elements of medical imaging simulation and the way to integrate all the outputs together. Each time data of 10GB, 20GB, 30 GB and then 40GB can been processed respectively and the focus is to identify how long data fusion (forming skeleton) may take since the emphasis is on data fusion to complete medical imaging simulation. Thus, the completion time of M-Fusion and M-Update have been measured when the data size varies.

Figure 9 (a) shows the completion time for executing of M-Fusion and M-Update for 10 GB of data between 10 and 100 nodes. All the completion time drops as an inverse exponential curve when the number of nodes has increased up to 100 nodes. Both functions stay close with each other in their completion time. Similarly, Figure 9 (b), 9(c) and 9(d) shows the completion time for executing of M-Fusion and M-Update for 20 GB, 30GB and 40GB of data respectively between 10 and 100 nodes. They have similar inverse exponential curve shape, except the slope has become more gentle when the size of data increases. In other words, the impact of running M-Fusion and M-Update on the completion time has decreased when the size of data increases. All the completion time can be done in less than 600 seconds for 40GB of data. Additionally, the impacts to the completion time with the increased nodes are not significant.

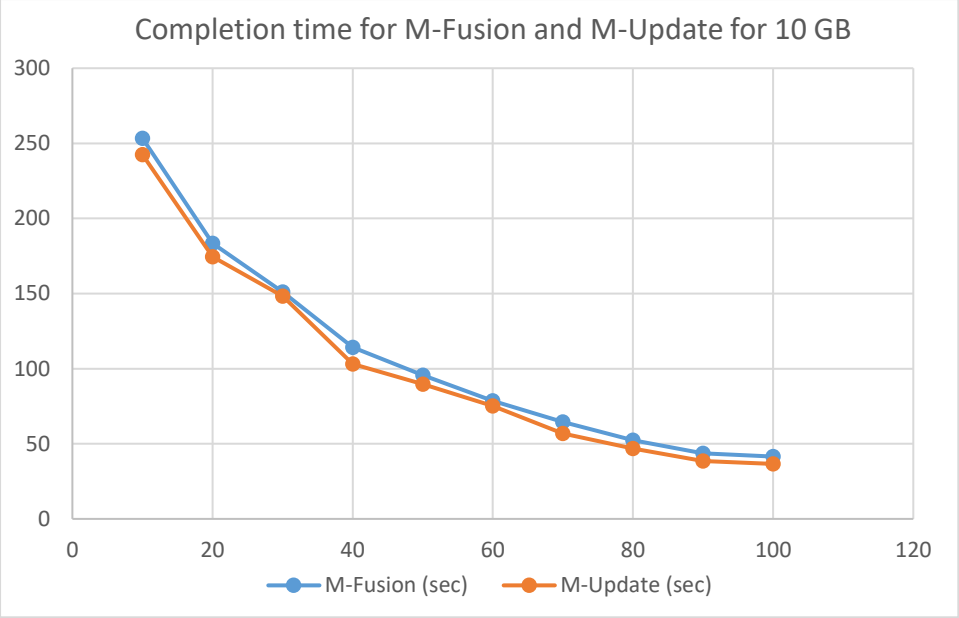


Fig. 7 (a): Completion time for M-Fusion and M-Update for 10 GB

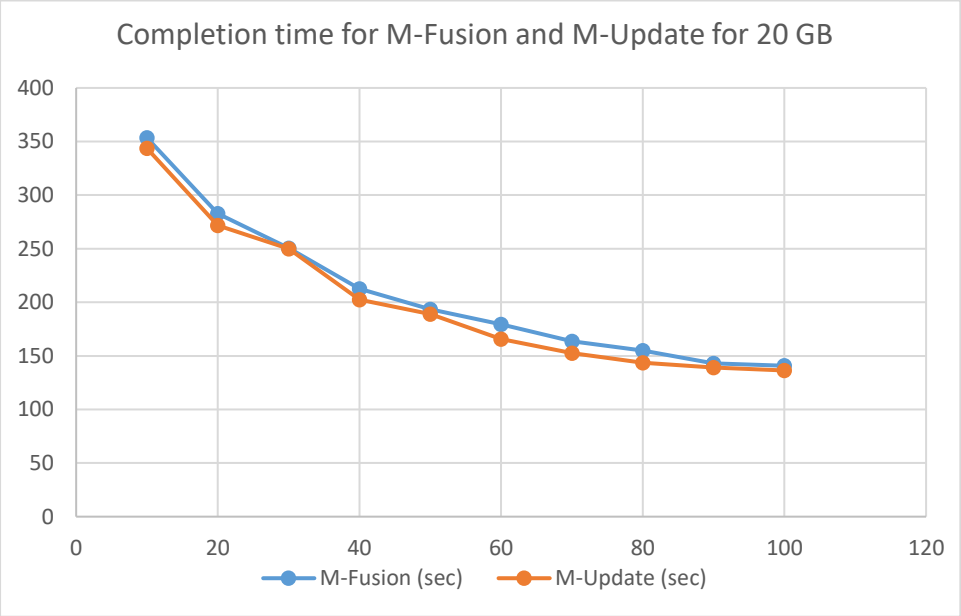


Fig. 7 (b): Completion time for M-Fusion and M-Update for 20 GB

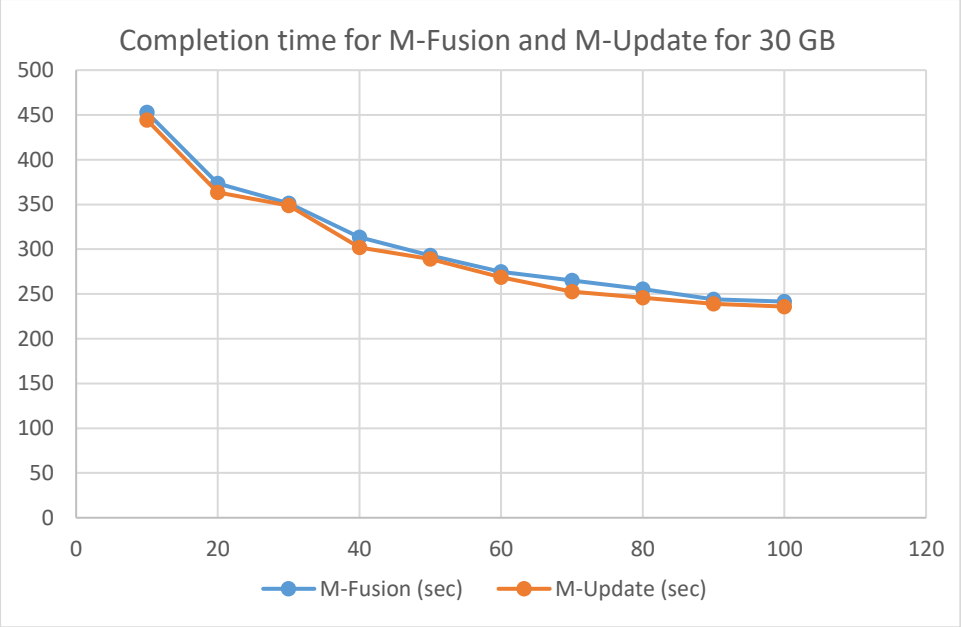


Fig. 7 (c): Completion time for M-Fusion and M-Update for 30 GB

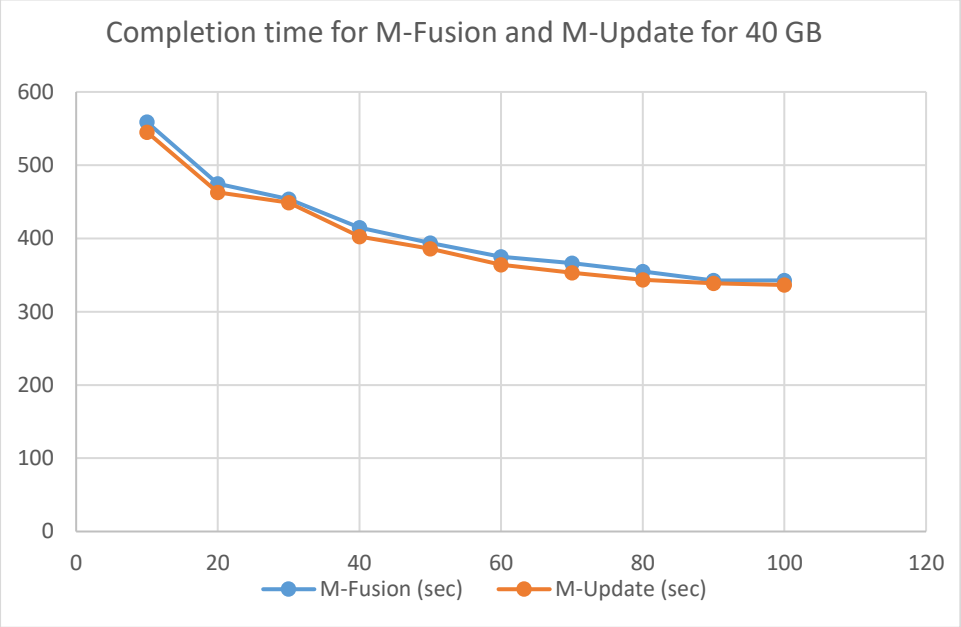


Fig. 7 (d): Completion time for M-Fusion and M-Update for 10 GB

5. Discussion

Advanced computational intelligence is useful to medical imaging simulations with the following reasons. First, costs of running very expensive experiments and medical equipment in laboratories can be reduced. Simulations can be done many times and outputs can be computed or queried at any time to support reproducibility of results [21-22]. Second, it provides a pioneering way to explore multimodal biomedical research by providing simulations similar to the real organs, such as rain segmentation to represent the brain cell intensities to respond to dancing. Third, there are areas that biomedical imaging cannot be easily adopted such as investigating at the micro-level such as genes and proteins, and investigating whether they have the liabilities of triggering the development of malignant tumors or cancerous cells. Forth, simulating biomedical work can be presented by

visualization and analytics, so that complex biological science and physiological science can be explained more easily with the visual aids.

Computational intelligence can combine the "beauty of computer and biological science" by analogizing the similarities between theories of digital surfaces with development of "skeletons", and improved MapReduce framework. Our research contributions are as follows. First, we develop fusion algorithm and MapReduce to demonstrate how to simulate medical imaging. Second, we have demonstrated how to perform simulation to investigate genes that are prone to trigger cancers and perform inspection. This technique cannot be easily achieved by the use of medical imaging alone which can provide a good alternative to medical imaging. Third, the completion time for all data fusion can be undertaken all under 600 seconds for processing up to 40 GB of data, with a good performance evaluation achieved.

6. Conclusion and Future Work

This paper demonstrates proofs-of-concept of simulating medical imaging by our advanced computational intelligence technique. We identify the similarities between the digital surface theories and our MapReduce framework with fusion algorithm. We can regard each biological unit as a data and the treat them the same way to process and fuse data together. We have explained our MapReduce functions and fusion algorithm, in which M-Fusion and M-Update can get the outputs of P-Reduce functions together. This allows simulating small units into clusters of units, and eventually the entire simulated medical image. Our proposed technique has the advantages than medical imaging alone, by simulating genes, proteins and immunity that cannot be easily be achieved by medical imaging alone. Examples of BIRC3, BIRC6, CCL4 and KLKB1 have been demonstrated. Inspection on genes to check any signs of cancers can be performed. Medical imaging on brain segmentation has also been explained. Additionally, performance evaluation of our fusion algorithm: M-Fusion and M-Update was undertaken that all 10GB, 20GB, 30 GB and 40GB of data had completion time between 45 seconds and 580 seconds, between 10 and 100 nodes. Therefore, our wok has demonstrated a cost-effective, useful and effective way of developing analytics and visualization to influence biomedical imaging simulation. Our future work will include more varieties of gene simulation and enhanced fusion algorithm to take on 100 GB of data analysis and fusion.

7. References

[1] Leblond, F., Davis, S. C., Valdés, P. A., & Pogue, B. W. (2010). Pre-clinical whole-body fluorescence imaging: Review of instruments, methods and applications. Journal of photochemistry and photobiology B: Biology, 98(1), 77-94.

[2] Weber, G. W. (2015). Virtual anthropology. American journal of physical anthropology, 156(S59), 22-42.

[3] Rose, N. S., & Abi-Rached, J. M. (2013). Neuro: The new brain sciences and the management of the mind. Princeton University Press.

[4] Pal, S. N., Duncombe, C., Falzon, D., & Olsson, S. (2013). WHO strategy for collecting safety data in public health programmes: complementing spontaneous reporting systems. Drug safety, 36(2), 75-81.

[5] Civera, J., Grasa, O. G., Davison, A. J., & Montiel, J. M. M. (2010). 1-Point RANSAC for extended Kalman filtering: Application to real-time structure from motion and visual odometry. Journal of Field Robotics, 27(5), 609-631.

[6] Kennedy, J. (2006). Swarm intelligence. In Handbook of nature-inspired and innovative computing (pp. 187-219). Springer US.

[7] Kim, C. E. (1983). Three-dimensional digital line segments. IEEE Transactions on Pattern Analysis and Machine Intelligence, (2), 231-234.

[8] Chen, L., & Zhang, J. (1993, June). Digital manifolds: an intuitive definition and some properties. In Proceedings on the second ACM symposium on Solid modeling and applications (pp. 459-460). ACM.

[9] Malgouyres, R. (1997). A definition of surfaces of Z 3 A new 3D discrete Jordan theorem. Theoretical Computer Science, 186(1), 1-41.

[10] Bertrand, G., & Malgouyres, R. (1999). Some topological properties of surfaces in Z3. Journal of Mathematical Imaging and Vision, 11(3), 207-221.

[11] Brimkov, V. E., & Klette, R. (2004, January). Curves, hypersurfaces, and good pairs of adjacency relations. In IWCIA (pp. 276-290).

[12] Klette, R., & Rosenfeld, A. (2004). Digital geometry: Geometric methods for digital picture analysis. Elsevier.

[13] Klette, G., & Pan, M. (2004, November). 3D topological thinning by identifying non-simple voxels. In IWCIA (pp. 164-175).

[14] Klette, G. (2006). Branch voxels and junctions in 3D skeletons. Combinatorial Image Analysis, 34-44.

[15] Chang, V. (2017). Towards data analysis for weather cloud computing. Knowledge-Based Systems, 127, 29-45.

[16] Lopez, L. M., Bastin, M. E., Maniega, S. M., Penke, L., Davies, G., Christoforou, A., ... & Porteous, D. J. (2012). A genome-wide search for genetic influences and biological pathways related to the brain's white matter integrity. Neurobiology of aging, 33(8), 1847-e1.

[17] Turner, N., & Grose, R. (2010). Fibroblast growth factor signalling: from development to cancer. Nature reviews. Cancer, 10(2), 116.

[18] Wong, R. S. (2011). Apoptosis in cancer: from pathogenesis to treatment. Journal of Experimental & Clinical Cancer Research, 30(1), 87.

[19] Balkwill, F. (2004). Cancer and the chemokine network. Nature reviews. Cancer, 4(7), 540.

[20] Moreau, M. E., Garbacki, N., Molinaro, G., Brown, N. J., Marceau, F., & Adam, A. (2005). The kallikrein-kinin system: current and future pharmacological targets. Journal of pharmacological sciences, 99(1), 6-38.

[21] Gentleman, R. (2005). Reproducible research: A bioinformatics case study. Statistical applications in genetics and molecular biology, 4(1), 1034.

[22] Huang, D., Tory, M., Aseniero, B. A., Bartram, L., Bateman, S., Carpendale, S., ... & Woodbury, R. (2015). Personal visualization and personal visual analytics. IEEE Transactions on Visualization and Computer Graphics, 21(3), 420-433.